Association for Information Systems
AIS Electronic Library (AISeL)

WHICEB 2018 Proceedings

Wuhan International Conference on e-Business

Summer 6-30-2018

Network Structure Mining and Evolution Analysis - Based on BA Scale-Free Network Model

Rui chen
School of Business Administration, Zhongnan University of Economics and Law, 430073, Wuhan, China

Ting He

Network Structure Mining and Evolution Analysis - Based on BA Scale-Free Network Model

Rui chen[1] *,Ting He*
School of Business Administration,
Zhongnan University of Economics and Law,
430073, Wuhan, China

Abstracts: The massive adoption of the Internet facilitates growth of online social networks, in which information can be exchanged in a more efficient way. Such as products, user accounts, web pages, there may be a variety of objects suitable to structurize this kind of networks. As a result, this gives the networks complexity and dynamics. The work in this paper is aiming to studying the topological property of online social network structure from the aspect of dynamics, and make clear the evolution processes of the networks. This is done by a Mean-Field analysis of network growth based on BA Scale-Free network model. Data resources come from the Chinese online e-commerce platform you.163.com and graphs are modeled through commentator and mutual comments by calculating degree distribution of the networks. We build a growing random model for forecasting dynamics of degree evolution. Finally, we use data set on Sina Weibo to test the model and the results are satisfying.

Key words: social network; degree distribution; BA model

1. INTRODUCTION

Graphs provide a useful abstract for modeling various networks. Recently this approach has been employed intensively in online business researches. Subscribers, products and even web pages can be modeled as nodes where followship, comments and scores can be links [1]. Online social networks are obviously suitable for graph model and therefore the network structure characteristics can be studied. With the help of large scale computing platform and the appropriate data mining techniques, analysis from aspect of network science will give deeper insight into online social networks.

Node degrees may be one of the most basic and important topology properties for a graph [2]. Instead of counting statical node's degrees, we observe the over-all distribution of the degrees. From the perspective of statistical probability, we denote $p(k)$ as the proportion of nodes which has degree k, and the p is exactly the distribution of nodes' degree we familiar with. The distribution of nodes' degree gives us a straightforward indication of how the whole network is linked. Furthermore, different distribution of degree indicates different network properties. The distribution of ER Random-Network-Model proposed by Erdős in 1959 obeys Gaussian distribution [3]. In 1999, Faloutsos et al found that the degree distribution of the Internet has a strong power law distribution [4]. In many recent studies, it was found that the distribution of the degree of social networks and that of time intervals of human behavior approximate obeys the power law distribution [5]. As the Internet data resources are increasing dramatically, the degree distribution of network shows more diversified distribution forms.

The research of this paper is based on BA Scale-Free network model [6]. BA Scale-Free network model is a

[1] Rui chen
School of Business Administration
Zhongnan University of Economics and Law
430073, Wuhan, China
Email：ruich@hotmail.com

network model based on node growth choice and power law distribution, which is an important random model outside the ER stochastic network. In the BA model, the edge probability is proportional to the size of the node itself, which reproduces the growth process of some networks [6]. The main feature of the BA Scale-Free model is that the probability of the node selecting the link varies between the two extremes, one is a new node uniform random selection of the original, the other the node is selected according to the current degree of the existing node, which called preferred attachment by Erdős and Albert [7]. We mainly consider the hybrid model and optimize it. Through empirical analysis, the structure of this model not only reflects the correlation of degree, but also shows the basic same type characteristics, which is also consistent with many social networks observed. When the process contains some form of preferred attachment, we also see that some large axis nodes appear in the network, which produces a smaller diameter than in the Poisson random network [8].

2. BA SCALE-FREE NETWORK MODEL AND ITS CHARACTERISTIC ANALYSIS

BA model, proposed by R. Albert and A. Barabasi, is a Scale-Free model, which contains two assumptions. i.e.,

Assumption 1: Networks are growing by time. Assuming a network initially has m0 nodes, a new node will be added in each time unit, and it will be connected to m of original nodes（$m \leq m_0$）. In this case we call it uniformly random choice link [8]. The final distribution of degree of nodes is:

$$F_t(d) = 1 - e^{\frac{d-m}{m}} \tag{1}$$

Assumption 2: The probability of an existing node to be connected to a new node increases proportionally with the degree of it, that is to say, for an existing node i, the probability of connecting to the new born node is m times the degree of node i to the total degree of all existing nodes at time t. i.e.:

$$m\frac{d_i(t)}{\sum_{j=1}^{t} d_j(t)} \tag{2}$$

In this case we call it preferred attachment [9].The final cumulative distribution function is:

$$F_t(d) = 1 - m^2 d^{-2} \tag{3}$$

Both of the above case are extreme cases. One is that a new born node uniformly chooses some of existing nodes to connect to. The other is that the new born node selects existing nodes to connect to according to current degree of the existing nodes. The main feature of the BA scale-free model is that the probability that a born node chooses existing node to link varies between the two extreme cases mentioned above. That is, a new node is connected by two different processes: a uniformly random link and a preferred connecting. Each new node forms m links, in which the probability it uniformly-randomly connect to original nodes is α, and the probability it connect to existing nodes with preferences is $1-\alpha$. The probability distribution function for this mixed model is:

$$F_t(d) = 1 - \left(\frac{m + \frac{2\alpha m}{1-\alpha}}{d + \frac{2\alpha m}{1-\alpha}}\right)^{\frac{2}{1-\alpha}} \tag{4}$$

When $\alpha = 0$, the distribution function is $F_t(d) = 1 - m^2 d^{-2}$, which is completely under the condition of preferred selection, that is, the same as (3); When $\alpha \to 1$, its distribution function approximately equals to $1 - e^{(d-m)/m}$, which is close to the probability distribution (1).

3. THE CALCULATION OF BA SCALE-FREE NETWORK HYBRID MODEL EXTENSION

The previous model usually has a single node at each time. If a fixed node is born in each period, the characteristics of these systems are generally unchanged. However, if the number of newborn nodes grows with time, the degree distribution will change. Consider an extension of the mixed model such that the number of nodes born during each period grows over time. Assuming that the number of new nodes born at time t is gn_t, where nt is the number of existing nodes at time t, and $g > 0$ is the growth rate. We estimate the distribution of degrees using continuous time approximations of degree distributions [8,9].

3.1 Continuous time approximations of degree distributions.

Given nodes numbered with $i \in (1,2,3,\dots,t,\dots)$, considering a growing network, each node is denoted by its born time, then the degree of node i born at time t can be represented as:

$$d_i(t) = \varphi_t(i) \tag{5}$$

Where $\varphi^{-1}(d)$ represents the number of the node with degree d, and which is the inverse of $d_i(t)$ .While the number is set strictly by $(1,2,3,\dots,t,\dots)$, that is also the number(count) of nodes of which the degrees are larger than d, so the proportion of nodes whose degree is less than d is:

$$\frac{t-\varphi^{-1}(d)}{t} = 1 - \frac{\varphi^{-1}(d)}{t} \tag{6}$$

That is, the distribution function is:

$$F_t(d) = 1 - \frac{\varphi^{-1}(d)}{t} \tag{7}$$

Considering the new born nodes added at time t is gn_t, these nodes randomly choose m of the t exist nodes to connect with probability α, thus, the initial condition of node i is $d_i(i) = m$ and for $t > i$ the change of degree over time is approximately:

$$\frac{dd_i(t)}{dt} = \frac{\alpha m g n_t}{n_t} \tag{8}$$

These nodes selected m of the t existing nodes with preference with probability $1-\alpha$, then the initial condition of node i is $d_i(i) = m$ and for $t > i$ he change of degree over time is approximately:

$$\frac{dd_i(t)}{dt} = \frac{(1-\alpha) m g n_t d_i(t)}{2 n_t m} \tag{9}$$

So the node’s change of degree over time can be represented as:

$$\frac{dd_i(t)}{dt} = \left[\frac{\alpha m}{n_t} + \frac{(1-\alpha) m d_i(t)}{2 n_t m}\right] g n_t$$

$$\frac{dd_i(t)}{dt} = \frac{2\alpha m g + (1-\alpha) g d_i(t)}{2} \tag{10}$$

Where $d_i(i) = m$. The final result is:

$$d_i(t) = \frac{m(1+\alpha)e^{\frac{g(1-\alpha)(t-i)}{2}} - 2\alpha m}{1-\alpha} = \varphi_t(i) \tag{11}$$

That is:

$$\varphi_t^{-1}(d) = t - \frac{2\ln\left[\frac{(1-\alpha)d + 2\alpha m}{m(1+\alpha)}\right]}{g(1-\alpha)} \tag{12}$$

And the distribution function is:

$$F_t(d) = 1 - \frac{t}{n_t} + \frac{2\ln\left[\frac{(1-\alpha)d + 2\alpha m}{m(1+\alpha)}\right]}{gt(1-\alpha)n_t} \tag{13}$$

Where

$$n_t = m(1+g)^t \tag{14}$$

And the final distribution of degrees is:

$$F_t(d) = 1 - \frac{t}{m(1+g)^t} + \frac{2\ln\left[\frac{(1-\alpha)d + 2\alpha m}{m(1+\alpha)}\right]}{mgt(1-\alpha)(1+g)^t} \tag{15}$$

4. EMPIRICAL ANALYSIS

4.1 Basic attribute characteristics of comment- interactive network

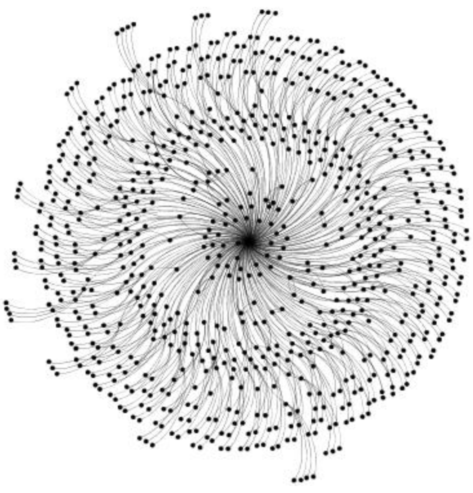
We have obtained the comments on the topic "black pig" of the microblogs of “you.163.com” from March 28, 2017 to April 2, 2017 (as shown in table 1.). Consider the degree distribution of the commenter network, where the link represents the response between the two commentators. The final distribution is shown in Table2, the final Network topology is shown in Figure1.In addition, we can also see from Figure1. that you.163.com network‘s degree distribution obeys power-law distribution, and the network shows strong heterogeneity.

Table1. you.163.com’s basic properties of the network structure

Time (t)	3.28	3.29	3.3	3.31	4.1	4.2	4.3
Node(n)	468	775	939	969	990	996	999
Edge (e)	480	823	1049	1088	1109	1115	1119
Average degree (d)	2.051	2.124	2.234	2.243	2.24	2.239	2.24

Table2. Frequency distribution of the interactive network of reviewers

degree	1	2	3	4	5	6	7	8	9	11	12	17	25	29	393
amount	631	29	262	47	17	6	4	1	2	1	1	2	1	1	1



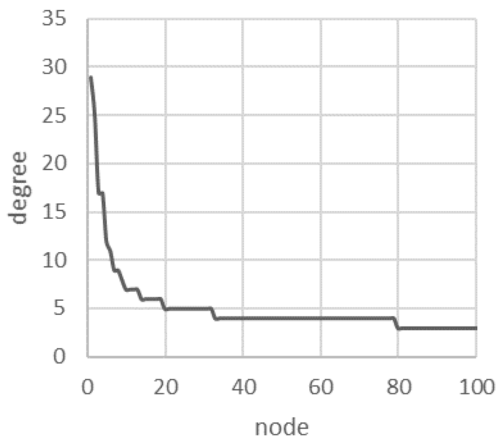


Figure 1. you.163.com’s network topology graph and degree distribution

4.2 Degree correlation

Consider a growing hybrid random network formation process as described in Section3.1. Under the mean-field estimate, a node i’s degree is larger than a node j’s degree at time t after both are born if and only if i is older than j [10]. A number of social networks have positive correlation in their degree distribution [11], Figure 2. shows you.163.com’s network evolution diagram in 3 units of time, from the figure we can clearly see the degree of positive correlation. As time goes on, the degree of the original nodes obviously increases, which further reflects that older and higher nodes grow faster than young and lower nodes, and this richer and richer process results in Scale -Free distribution.

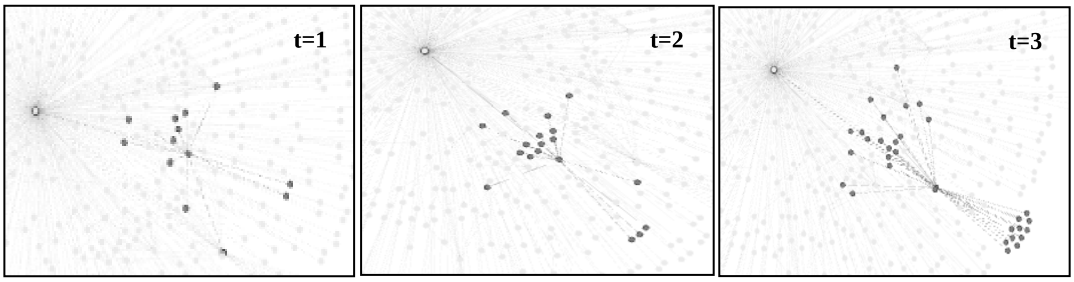


Figure 2. you.163.com's network evolution diagram

4.3 Fitting hybrid degree distributions to data

First, we calculate m directly, since m is the number of connections formed during each period, so it is half of the increase in each period. The total degree is $2tm$, so m is half of the average [12]. The mean of the network is 2.24, so m is about 1.12, and according to table 1, g is calculated as 15.5%. The initial guess of α, α0 is used as a starting point, the fitting of the estimation is investigated, and the fixed point of the process is investigated to estimate the parameter α [13]. We calculate α according to table 1 and table 2, as shown in table 3. In this case, the final estimated α is approximately 0.823.

Table3. Initial parameter estimate α0 and final fit α1

α0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	0.99	0.999
α1	-0.63	-0.4	0.34	0.27	0.38	0.55	0.63	0.823	0.825	0.86	0.883

5. CONCLUSIONS

The process of random graph is often complicated, especially when nodes randomly enter into a link and have different degrees of distribution. Based on the BA model, this paper studied the growing random network model and illustrated some of its characteristics. What is important is that its result is more robust than Poisson random network, and in extreme cases, it provides an explanation for scale-free distribution. In order to overcome the challenges of social media data mining, this study provides a structured mechanism to extract values from the data. We can see that there is a significant characteristic of the teletype network which is the heterogeneity, the degree is positively correlated, and the degree distribution is the power law distribution. These empirical results are also consistent with the empirical results of some previous e-commerce networks. However, if you add more network topology considerations, the effect will be more significant. A deep understanding of network structure allows us to develop more effective strategies to serve market decisions.

REFERENCES

[1] Jiang J, Wilson C, Wang X, et al. Understanding latent interactions in online social networks[J]. ACM Transactions on the Web (TWEB), 2013, 7(4):18.
[2] Albert R. Barab'asi: Statistical mechanics of complex networks[J]. Rev.mod.phys, 2002, 74(1):2002.
[3] Barabási A, Albert R. Emergence of Scaling in Random Networks[J]. Science, 1999, 286(5439):509.

[4] Albert R, Jeong H. Internet: Diameter of the World-Wide Web[J]. Nature, 1999, 401(6):130-131.

[5] M. Faloutsos, P. Faloutsos, C. Faloutsos. On power-law relationships of the internet topology[C].Proceedings of the Conference on Applications, Technologies, Architectures, and Protocols for Computer Communication, Cambridge, MA, USA, 1999, 251–262

[6] Barabasi A L, Albert R. Emergence of scaling in random networks[J]. Science, 1999, 286(5439):509.

[7] Jeong H, Tombor B, Albert R, et al. The large-scale organization of metabolic networks[J]. Nature, 2000, 407(6804):651-4.

[8] Bernard H R, Johnsen E C, Killworth P D, et al. Comparing four different methods for measuring personal social networks [J]. Social Networks, 1990, 12(3):179-215.

[9] Barabási A L, Albert R, Jeong H. Scale-free characteristics of random networks: the topology of the world-wide web[J]. Physica A Statistical Mechanics & Its Applications, 2000, 281(1–4):69-77.

[10] Jackson M O, Rogers B W. Meeting Strangers and Friends of Friends: How Random Are Social Networks?[J]. American Economic Review, 2007, 97(3):890-915.

[11] Littman M L, Kearns M, Singh S. An Efficient, Exact Algorithm for Solving Tree-Structured Graphical Games[J]. Advances in Neural Information Processing Systems, 2001:817-823.

[12] Besag, J. E. (1974) “Spatial interaction and the statistical analysis of lattice systems (with discussion),”Journal of the Royal Statistical Society, Series B, 36:196-236.

[13] Kakade S, Kearns M, Langford J, et al. Correlated equilibria in graphical games[C]// ACM Conference on Electronic Commerce. ACM, 2003:42-47.